

INPUT FOR THE CRU REVIEW: M J Kelly 25.III.10

Comments on Briffa Papers (nos 2-6 on list)

Initial Response after First Reading: (will reread and comment on each)

All papers are involved with trying to extract past climate information from tree-ring data. There are two stages in this, the first trying to take the raw data and remove features that have their origin outside what is known or thought to be relevant, such as that older trees tend to grow more slowly. At this stage the choice of initial data accepted is also important. Because of other factors (precipitation, hours of sunlight, attitude (north facing etc)) all have a bearing on tree growth whatever the climate, trees are used only from high latitudes and near the tree-line where any actual climate dependence is likely to be more prominent. The second stage is to try to extract climatic inferences from this suitably prepared input data.

My overriding impression that this is a continuing and valiant attempt via a variety of statistical methods to find possible signals in very noisy and patchy data when several confounding factors may be at play in varying ways throughout the data. It would take an expert in statistics to comment on the appropriateness of the various techniques as they are used. The descriptions are couched within an internal language of dendrochronology, and require some patience to try and understand.

There is no evidence, as far as I am concerned, of anything other than a straightforward scientific exercise within the confines described above. The papers are full of suitable qualifications about the limitations of the data and the strength of the inferences to be drawn from them. I find no evidence of blatant mal-practice. That is not to say that, working within the current paradigm, choices of data and analysis approach might be made in order to strain to get more out of the data than a dispassionate analysis might permit.

There are however some more detailed qualifications:

(i) I take real exception to having simulation runs described as experiments (without at least the qualification of 'computer' experiments). It does a disservice to centuries of real experimentation and allows simulations output to be considered as real data. This last is a very serious matter, as it can lead to the idea that real 'real data' might be wrong simply because it disagrees with the models! That is turning centuries of science on its head.

(ii) The reading of the papers was made rather harder by the quality of the diagrams, and the description of the vertical axes on a number of graphs. When numbers on the vertical axis go from -2 to +2 without being explicitly labelled as percentage deviations, temperature excursions, or scaled correlation coefficients, there is potential for confusion.

(iii) I think it is easy to see how peer review within tight networks can allow new orthodoxies to appear and get established that would not happen if papers were written for and peer-reviewed by a wider audience. I have seen it happen elsewhere. This finding may indeed be an important outcome of the present review.

More detailed comments on the Briffa papers, by paper, on a second reading:

(2) 'Reduced sensitivity of recent tree-growth to temperatures at high northern latitudes', K R Briffa et al, *Nature* 391 678-82 (1998)

This is a short contribution to the 'divergence debate'. Samples are taken from 300 sites, although there is no hard-and-fast rule that is used to discriminate what is and what is not included. A range of statistical analyses are done, with particular correlations ranging from 34% to 85% (with average 60%) in the period 1881-1960 which drop by about 20% when the period is extended to 1881-1981. To an untrained eye, the raw data is very noisy, and even then the raw data has been detrended of age-trends in individual trees, and the subsequent data is scaled to have zero average and unit variance over the time period before being plotted. This means that correlations can only be qualitative and temporal. A variety of suggestions are made for the growing divergence of the tree-ring and the instrumental record over the last 50 years, each of which could be convolved in the data further back, but no one thing is concluded to be the primary cause. While it may be a laudable intent to make these correlations, it would be easy to remain sceptical as to their real value, and especially if one tried to make and insist upon quantitative conclusions.

(3) 'Trees tell of past climates: but are they speaking less clearly today', K R Briffa et al, *Phil. Trans. Roy. Soc. London. B* 353 65-73 (1998)

This is a longer version of the previous paper. 'Inferring the details of past climate variability from tree-ring data remains a largely empirical exercise, but one that goes hand-in-hand with the development of techniques that seek to identify and isolate the confounding influence of local and larger-scale non-climatic factors.' Figure 2 shows dramatic differences in long time-scale temperature information reconstructed from the same tree-ring data using two different techniques to removed localized age biases – they differ by a factor of five in scale! Because the one with the larger excursion retains greater long-time-scale changes, (e.g. the medieval warm period and the little ice age) it is regarded as superior. I remain worried about how the actual absolute scale of temperature excursion is decided upon, as shown in Figure 3. Figure 5 has no vertical axis description: it says it is a plot of standardised anomalies, but it has an average of -0.3 and a standard deviation of 1.1, but what? Section 5 raises the 'divergence' issue. Section 6 looks at basal area increments and maximum density, showing that the former rises linearly from 1850 to 1950 and flattens, while the latter is flat from 1850 to 1950 and then falls. It is hard directly to correlate this aspect with the anthropogenic hypothesis of climate warming. Some features do correlate – others don't – so where is the rigorous test of the significance of correlation or lack of it?

(4) 'Annual climate variability in the Holocene: interpreting the message of ancient trees', K R Briffa, *Quaternary Sciences Reviews* 19 87-105 (2000)

This is a major paper reviewing and updating his work over the 1990s. Referring to dendroclimatology supporting the notion that the last 100 years have been unusually warm in the context of the last 2000 years, Briffa says: ‘However, this evidence should not be considered unequivocal.’ He also states ‘The interrelationships between large-scale patterns of temperature, precipitation and atmospheric pressure variability also mean that networks of climate sensitive tree-ring chronologies can be used to make statistical inferences about the past behaviour of circulation patterns or important circulation indices.’ The Figure 1 shows several selected reconstructions of summer temperatures over the last 2000 years. I am not sure just how the vertical scale (temperature) is calibrated, other perhaps (but not stated explicitly) than by correlation with the recent instrument record. I have trouble with the vertical axis of Figure 3, relating to moisture reconstructions. The major sections 3 and 4 of this paper work to reconstruct the major circulation patterns in the northern and southern hemisphere in so far as this can be done from tree-ring data. In terms of a chronology of events (e.g. volcanic eruptions) there are some correlations, but the actual excursions of temperature etc are less convincing. He points out the need for more data from the Himalayas and other regions. He also points out that the 20th century data seems anomalous, and speculates on what is happening, but does not conclude why it is happening.

(5) ‘Low-frequency temperature variations from a northern tree-ring density set’ K R Briffa et al, *Journal of Geophysical Research* 106 2929-41 (2001)

This paper uses a new statistical technique ‘age band decomposition’ to examine northern hemisphere climate change over the last 600 years with the intent of preserving some of the longer-timescale variability that is lost by other techniques. The reconstruction results in generally lower temperatures for earlier times, notably the 17th century, but the northern Siberia had 15th century summers warmer than those in the 20th century. The Figure 1 shows the full gamut of raw data which is described as climate signal + age signal + noise, and what happens when all the data from trees that are 21-40 and 51-70 years old are averaged, and then combined. This is yet another technique for detecting a weak signal in noisy and patchy data. Plate 2 contains averaged data from 9 different regions, and there is really not much inter-correlation signifying either short events or multidecadal events. Further on, plate 4 shows a range of reconstructions compatible with the same input data, and while results from 1700 to 1950 look mutually consistent, the results before then or after are certainly not. Their plate 3 is an often quoted diagram of six large-scale reconstructions, with a standard deviation of 0.1C variability at 1900, increasing to 0.3C at 1700.

(6) ‘Trends in recent temperature and radial tree growth spanning 2000 years across northwest Eurasia’, K R Briffa et al, *Phil. Trans. R. Soc. B* 363 2271-84 (2008).

The first sentence of the text refers to climate model experiments, which offends me! This more recent paper looks at regional reconstructions over the last 2000 years, showing strong regional variations. ‘A set of long tree-ring chronologies provides empirical evidence of

association between inter-annual tree growth and local, primarily summer, temperature variability at each location. These data show no evidence for the recent breakdown in this association as has been found at other high-latitude Northern Hemisphere locations.’ That means, there is no divergence here! Yet another technique, Kendall’s concordance, is used to ‘show strong evidence that the extent of recent widespread warming across northwest Eurasia, with respect to 100- to 200- year trends, is unprecedented in the last 2000 years.’ This involves data from three regions, Fennoscandia, Yamal and Avam-Taimyr. Many of the vertical scales are described as ‘index values’ so that the chronologies can show events but the absolute excursion amplitudes of any parameters are not calibrated. Figure 5 show that the various trend parameters and means show that the observations that are two or more standard deviations positive are mainly from 1900-1946. In section 5 it is shown how correlation plots between the regional curve standardized chronologies and (i) monthly mean temperatures over 1950-1994 and (ii) a sequence of temperatures averaged over successive periods of five days. In the final section, one reads: ‘These results are superficially consistent with the expected patterns of increasing high-latitude warming suggested by GCM simulations of possible future climates under enhanced atmospheric GHG emissions. However, a simple analysis of one such experiment, under natural and GHG forcing for the last 250 years, while showing consistently increasing concordance between simulated temperatures in the regions of our chronologies, failed to produce results that could be distinguished from the results of a similar experiment driven only with natural (i.e. non-anthropogenic) forcings.’ The line between positive conclusions and the null hypothesis is very fine in my book.

Comments on Jones Papers (7-9) 25.III.10

(7) 'Hemispheric and Large-scale surface air temperature variations: an extensive revision and an update to 2001', P D Jones et A Moberg, *Journal of Climate* 16 209-223 (2003)

The title describes the contents. Section 2 focuses on data, section 3 on interpolation onto a grid, section 4 analyses the land data and section 5 looks at combined marine and land data. I worry about the sheer range and the ad hoc/subjective nature of all the adjustments, homogenisations etc of the raw data from different places. If Australia changes its way of calculating average temperature (from the average of max and min daily temperatures to a hourly or three-hourly average of the data) and get a -0.2C change, how representative is that change over the times before and after the switch in method of calculation? What if some of the eliminated outliers are genuine? There is plenty of openness about the limitations of the data. There is no evidence of overt scientific malpractice. That is not to absolve the authors of conscious or unconscious bias in making all the choices referred to above.

(8) 'Northern Hemisphere surface air temperature variations: 1851-1984', P D Jones et al, *Journal of Climate and Applied Meteorology* 25, 161-179 (1986a)

An attempt to get a database of 5(lat)x10(long) gridded temperature time series for the Northern Hemisphere over the period given. A long section 2 deals with inhomogeneity in the data and changes in the way data is calculated and presented, and urban heat island effects. Section 3 assesses the homogeneity of the data, and 4 presents the homogeneity results. Section 5 grids the temperatures data, 6 compares the results with other sources, 7 is concerned with incomplete data in earlier years, and 8 draws conclusions. All this happens before the latest concerns about rising temperature, so the main point of note was that 1921-1984 was 0.4C warmer than 1851-1920!

(9) 'Southern Hemisphere surface air temperature variations: 1851-1984', P D Jones et al, *Journal of Climate and Applied Meteorology* 25, 1213-1230 (1986b)

An attempt to get a database of 5(lat)x10(long) gridded temperature time series for the Southern Hemisphere over the period given, a companion and complement to the previous paper. Section 2 deals with the previous work, which is scarce and not as well characterised as for the NH. The section 3 deals with the data, its homogenization and gridding, section 4 discusses the effects of incomplete data. Section 5 deals with the results under headings such as comparisons with other temperature estimates, high-latitude and low latitude links, inter-hemisphere comparisons and temperature trends. Section 6 concludes. I am concerned about section 4: only 27% of the area is covered by land or adjacent land. Then there are correlations within models by selecting subsets of data showing downward trends as the 'distance' in time increases. I would be surprised at anything else. The handling of Antarctica is crude. Section 5c points out a number of correlations, and concludes that fluctuations in the NHT data do not need to be heeded too much. Even though only a few months later in submission there is a big change in emphasis on the global warming implications, showing no hint of significant cooling anywhere in the southern hemisphere. In neither of these papers is there any overt malpractice, but one can't eliminate the possibility

of conscious or unconscious bias in the choices of data. I just do wonder if a different hypothesis was being tested whether the same approach could give a very different answer.

Subsequent thoughts:

(1) My second reading reinforces my initial observations and concerns.

(2) On a personal note, I chose to study the theory of condensed matter physics, as opposed to cosmology, precisely on the grounds that I could systematically control and vary the boundary conditions of my object of study as an integral part of making advances. An elegant theory which does not fit good experimental data is a bad theory. Here the starting data is patchy and noisy, and the choices made are in part aesthetic, or designed to help a conclusion, rather than neutral. This all colours my attitude to the limited value of complex simulations that cannot be exhaustively tested against 'real' data from independent experiments that control all but one of the variables.

(3) Up to and throughout this exercise, I have remained puzzled how the real humility of the scientists in this area, as evident in their papers, including all these here, and the talks I have heard them give, is morphed into statements of confidence at the 95% level for public consumption through the IPCC process. This does not happen in other subjects of equal importance to humanity, e.g. energy futures or environmental degradation or resource depletion. I can only think it is the 'authority' appropriated by the IPCC itself that is the root cause.

(4) Our review takes place in a very febrile atmosphere. If we give a clean bill of health to what we regard as sound science without qualifying that very narrowly, we will be on the receiving end of justifiable criticism for exonerating what many people see as indefensible behaviour. Three of the five MIT scientists who commented in the week before Copenhagen on the leaked emails, (see <http://mitworld.mit.edu/video/730>) thought that they saw prima facie evidence of unprofessional activity.

(5) I think we should consider using the opportunity to make entirely positive recommendations that would improve the situation, such as (i) wider peer review to prevent narrow and premature orthodoxies being developed unchallenged and (ii) more effective engagement with the end-users of their findings beyond politicians and policy makers. Engineers seem more sceptical than others on the implications of the findings to date.

(6) There is late-breaking news about attempts to suborn the workings of the Journal of Geophysical Research, which I think we should examine and comment upon having heard from one of the co-authors before I was approached on this mission.

See <http://icecap.us/images/uploads/McLeanetalSPPIpaper2Z-March24.pdf>

MJK

My overall sympathy is with Ernest Rutherford: "If your experiment needs statistics, you ought to have done a better experiment."

Questions to Jones

- (1) How can we be reassured about the choice of which raw data from which stations are to be homogenised and then included in the gridded temperature data bases? Is there an algorithm that establishes the inclusion/exclusion of particular stations? If I were setting out to establish the lowest possible net temperature rise over the last century is consistent with the available data, what fraction of stations would then be included/excluded? Indeed, could the same data be 'coerced' to support a null hypothesis on global warming? Incidentally, how much lower could that temperature be?
- (2) What is a sceptical outsider to make of 'degrees of rigour of homogenisation' of the data, and also the variety of adjustments that have to be made on an ad hoc basis? How do you ensure that adjustment of adjustments do not introduce biases that are a significant fraction of the century temperature rise?
- (3) When updating database and redoing calculations, the scientific sceptics can point to adjustments of past data starting look like rewriting history (c.f. <http://wallstreetpit.com/20710-climategate-goes-back-to-1980>). How do you respond?
- (4) How does the initial formation and subsequent management of the various databases compare with best practice in general, and in the sector?
- (5) In presenting data and graphs, do you have a policy of always using the latest and best data, no matter what the message you are trying to convey? A 2006 Met Office diagram of central England Temperature, and not yet showing any turnover or turn down in 5-year averaged temperatures was used in an official report in 2009, when data showing the turn down was already available.
- (6) How, over time, have the overall results trended as more reliable data from Antarctica has been incorporated into the calculations? Has this incorporation made much difference?
- (7) Given that the outputs of your work are being used to promote the largest revolution mankind has ever contemplated, do you have any sense of the extent to which the quality control and rigour of approach must be of the highest standards in clear expectation of deep scrutiny?
- (8) Your critique of the paper by McLean, Freitas and Carter (2009) hinges on arcane aspects of statistical analysis, and they stand by their comments. I have recommended publication of data with a controversial explanation precisely to get the debate going. In other areas of science the best winds out by attrition: why not here?

Questions to: Briffa

(1) How can we be reassured about the choice of which raw data from which stations are to be selected, detrended and then included in the tree-ring data bases? Is there an algorithm that establishes the inclusion/exclusion? If I were setting out to establish the lowest possible net temperature rise over the last century is consistent with the available data, what fraction of tree-ring-data would then be included/excluded? Could I coerce the data to support a null hypothesis on global warming?

(2) In the range of papers we have reviewed, you have used a variety of statistical techniques in what is a heroic effort to get signals from noisy and patchy data. To what extent has this variety of techniques be reviewed and commented upon by the modern statistical community for their effectiveness, right use and possible weaknesses?

(3) Precisely how do you take a chronology and establish the actual amplitude of temperature excursions at a given time, especially at times that are outside the instrumental record.

(4) Do you think that if your papers had been regularly reviewed by a wider scientific community (i.e. outside dendrochronology) some of the current orthodoxies might have been tested more robustly? I am thinking of the comments raised by Gerd Burger in Science in 2007.

(5) What responsibility do you think that we, as a scientific community, have to ensure that the caveats in our papers are not glossed over by our scientific colleagues trying to formulate policy agendas?

(6) Have you had the opportunity to cross-correlate any of your findings with analogous studies of coral, giant claims, or any other temperature proxies? If so, what has emerged?

(7) Given that the outputs of your work are being used to promote the largest revolution mankind has ever contemplated, do you have any sense of the extent to which the quality control and rigour of approach must be of the highest standards in clear expectation of deep scrutiny?

Comments on Briffa Papers (nos 2-6 on list) 22.03.10

M J Kelly

Initial Response after First Reading: (will reread and comment on each)

All papers are involved with trying to extract past climate information from tree-ring data. There are two stages in this, the first trying to take the raw data and remove features that have their origin outside what is known or thought to be relevant, such as that older trees tend to grow more slowly. At this stage the choice of initial data accepted is also important. Because of other factors (precipitation, hours of sunlight, attitude (north facing etc)) all have a bearing on tree growth whatever the climate, trees are used only from high latitudes and near the tree-line where any actual climate dependence is likely to be more prominent. The second stage is to try to extract climatic inferences from this suitably prepared input data.

My overriding impression that this is a continuing and valiant attempt via a variety of statistical methods to find possible signals in very noisy and patchy data when several confounding factors may be at play in varying ways throughout the data. It would take an expert in statistics to comment on the appropriateness of the various techniques as they are used. The descriptions are couched within an internal language of dendrochronology, and require some patience to try and understand.

There is no evidence, as far as I am concerned, of anything other than a straightforward scientific exercise within the confines described above. The papers are full of suitable qualifications about the limitations of the data and the strength of the inferences to be drawn from them. I find no evidence of blatant mal-practice. That is not to say that, working within the current paradigm, choices of data and analysis approach might be made in order to strain to get more out of the data than a dispassionate analysis might permit.

There are however some more detailed qualifications:

(i) I take real exception to having simulation runs described as experiments (without at least the qualification of 'computer' experiments). It does a disservice to centuries of real experimentation and allows simulations output to be considered as real data. This last is a very serious matter, as it can lead to the idea that real 'real data' might be wrong simply because it disagrees with the models! That is turning centuries of science on its head.

(ii) The reading of the papers was made rather harder by the quality of the diagrams, and the description of the vertical axes on a number of graphs. When numbers on the vertical axis go from -2 to +2 without being explicitly labelled as percentage deviations, temperature excursions, or scaled correlation coefficients, there is potential for confusion.

(iii) I think it is easy to see how peer review within tight networks can allow new orthodoxies to appear and get established that would not happen if papers were written for and peer-reviewed by a wider audience. I have seen it happen elsewhere. This finding may indeed be an important outcome of the present review.

More detailed comments, by paper, on a second reading:

(2) 'Reduced sensitivity of recent tree-growth to temperatures at high northern latitudes', K R Briffa et al, Nature 391 678-82 (1998)

This is a short contribution to the 'divergence debate'. Samples are taken from 300 sites, although there is no hard-and-fast rule that is used to discriminate what is and what is not included. A range of statistical analyses are done, with particular correlations ranging from 34% to 85% (with average 60%) in the period 1881-1960 which drop by about 20% when the period is extended to 1881-1981. To an untrained eye, the raw data is very noisy, and even then the raw data has been detrended of age-trends in individual trees, and the subsequent data is scaled to have zero average and unit variance over the time period before being plotted. This means that correlations can only be qualitative and temporal.. A variety of suggestions are made for the growing divergence of the tree-ring and the instrumental record over the last 50 years, each of which could be convolved in the data further back, but no one thing is concluded to be the primary cause. While it may be a laudable intent to make these correlations, it would be easy to remain sceptical as to their real value, and especially if one tried to make quantitative conclusions.

(3) 'Trees tell of past climates: but are they speaking less clearly today', K R Briffa et al, Phil. Trans. Roy. Soc. London. B 353 65-73 (1998)

This is a longer version of the previous paper. 'Inferring the details of past climate variability from tree-ring data remains a largely empirical exercise, but one that goes hand-in-hand with the development of techniques that seek to identify and isolate the confounding influence of local and larger-scale non-climatic factors.' Figure 2 shows dramatic differences in long time-scale temperature information reconstructed from the same tree-ring data using two different techniques to removed localized age biases – they differ by a factor of five in scale! Because the one with the larger excursion retains greater long-time-scale changes, (e.g. the medieval warm period and the little ice age) it is regarded as superior. I remain worried about how the actual absolute scale of temperature excursion is decided upon, as shown in Figure 3. Figure 5 has no vertical axis description: it says it is a plot of standardised anomalies, but it has an average of -0.3 and a standard deviation of 1.1, but what? Section 5 raises the 'divergence' issue. Section 6 looks at basal area increments and maximum density, showing that the former rises linearly from 1850 to 1950 and flattens, while the latter is flat from 1850 to 1950 and then falls. It is hard directly to correlate this with the anthropogenic hypothesis of climate warming.

(4) 'Annual climate variability in the Holocene: interpreting the message of ancient trees', K R Briffa, Quaternary Sciences Reviews 19 87-105 (2000)

This is a major paper reviewing and updating his work over the 1990s. Referring to dendroclimatology supporting the notion that the last 100 years have been unusually warm in the context of the last 2000 years, Briffa says: 'However, this evidence should not be considered unequivocal.' He also states 'The interrelationships between large-scale patterns of temperature, precipitation and atmospheric pressure variability also mean that networks of climate sensitive tree-ring chronologies can be used to make statistical inferences about the past behaviour of circulation patterns or important circulation indices.' The Figure 1 shows several selected reconstructions of summer temperatures over the last 2000 years. I am not sure just how the vertical scale (temperature) is calibrated, other perhaps (but not stated explicitly) by correlation with the recent instrument record. I have trouble with the vertical axis of Figure 3, relating to moisture reconstructions. The major sections 3 and 4 of this paper work to reconstruct the major circulation patterns in the northern and southern hemisphere in so far as this can be done from tree-ring data. In terms of a chronology of events (e.g. volcanic eruptions) there are some correlations, but the actual excursions of temperature etc are less convincing. He points out the need for more data from the Himalayas and other regions. He also points out that the 20th century data seems anomalous, and speculates on what is happening, but does not conclude why it is happening.

(5) 'Low-frequency temperature variations from a northern tree-ring density set' K R Briffa et al, *Journal of Geophysical Research* 106 2929-41 (2001)

This paper uses a new statistical technique 'age band decomposition' to examine northern hemisphere climate change over the last 600 years with the intent of preserving some of the longer-timescale variability that is lost by other techniques. The reconstruction results in generally lower temperatures for earlier times, notably the 17th century, but the northern Siberia had 15th century summers warmer than those in the 20th century. The Figure 1 shows the full gamut of raw data which is described as climate signal + age signal + noise, and what happens when all the data from trees that are 21-40 and 51-70 years old are averaged, and then combined. This is another technique for detecting a weak signal in noisy and patchy data. Plate 2 contains averaged data from 9 different regions, and there is really not much inter-correlation signifying either short events or multidecadal events. Further on, plate 4 shows a range of reconstructions compatible with the same input data, and while results from 1700 to 1950 look mutually consistent, the results before then or after are certainly not. Their plate 3 is an often quoted diagram of six large-scale reconstructions, with a standard deviation of 0.1C variability at 1900, increasing to 0.3C at 1700.

(6) 'Trends in recent temperature and radial tree growth spanning 2000 years across northwest Eurasia', K R Briffa et al, *Phil. Trans. R. Soc. B* 363 2271-84 (2008).

The first sentence of the text refers to climate model experiments, which offends me! This more recent paper looks at regional reconstructions over the last 2000 years, showing strong regional variations. 'A set of long tree-ring chronologies provides empirical evidence of

association between inter-annual tree growth and local, primarily summer, temperature variability at each location. These data show no evidence for the recent breakdown in this association as has been found at other high-latitude Northern Hemisphere locations.’ That means, there is no divergence here! Yet another technique, Kendall’s concordance, is used to ‘show strong evidence that the extent of recent widespread warming across northwest Eurasia, with respect to 100- to 200- year trends, is unprecedented in the last 2000 years.’ This involves data from three regions, Fennoscandia, Yamal and Avam-Taimyr. Many of the vertical scales are described as ‘index values’ so that the chronologies can show events but the absolute excursion amplitudes of any parameters are not calibrated. Figure 5 show that the various trend parameters and means show that the observations that are two or more standard deviations positive are mainly from 1900-1946. In section 5 it is shown how correlation plots between the regional curve standardized chronologies and (i) monthly mean temperatures over 1950-1994 and (ii) a sequence of temperatures averaged over successive periods of five days. In the final section, one reads: ‘These results are superficially consistent with the expected patterns of increasing high-latitude warming suggested by GCM simulations of possible future climates under enhanced atmospheric GHG emissions. However, a simple analysis of one such experiment, under natural and GHG forcing for the last 250 years, while showing consistently increasing concordance between simulated temperatures in the regions of our chronologies, failed to produce results that could be distinguished from the results of a similar experiment driven only with natural (i.e. non-anthropogenic) forcings.’

Subsequent thoughts:

(1) My second reading reinforces my initial observations and concerns.

(2) On a personal note, I chose to study the theory of condensed matter physics, as opposed to cosmology, precisely on the grounds that I could systematically control and vary the boundary conditions of my object of study as an integral part of making advances. An elegant theory which does not fit the experimental data is a bad theory. This particularly colours my attitude to the limited value of complex simulations that cannot be exhaustively tested against real data from independent experiments that control all but one of the variables. Here the starting data is patchy and noisy and choices are in part aesthetic, or designed to help a conclusion, rather than neutral.

(3) Up to and throughout this exercise, I have remained puzzled how the real humility of the scientists in this area, as evident in their papers, including all these here, and the talks I have heard them give, is morphed into statements of confidence at the 95% level for public consumption through the IPCC process. This does not happen in other subjects of equal importance to humanity, e.g. energy futures or environmental degradation or resource depletion. I can only think it is the 'authority' appropriated by the IPCC itself that is the root cause.

(4) Our review takes place in a very febrile atmosphere. If we give a clean bill of health to what we regard as sound science without qualifying that very narrowly, we will be on the receiving end of justifiable criticism for exonerating what many people see as indefensible behaviour. Three of the five MIT scientists who commented in the week before Copenhagen on the leaked emails, (see <http://mitworld.mit.edu/video/730>) thought that they saw prima facie evidence of unprofessional activity.

(5) I think we should consider using the opportunity to make entirely positive recommendations that would improve the situation, such as (i) wider peer review to prevent narrow and premature orthodoxies being developed unchallenged and (ii) more effective engagement with the end-users of their findings beyond politicians and policy makers. Engineers are more sceptical than others on the implications of the findings to date.